(32)

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, A01K 67/027, 67/033, C12N 15/01 — A1

(11) International Publication Number: WO 97/44485

(43) International Publication Date: 27 November 1997 (27.11.97)

(21) International Application Number: PCT/GB97/01354

(22) International Filing Date: 16 May 1997 (16.05.97)

(30) Priority Data:
9610355.1 17 May 1996 (17.05.96) GB

(71) Applicant: HEXAGEN TECHNOLOGY LIMITED [GB/GB]; 214 Cambridge Science Park, Milton Road, Cambridge CB4 4WA (GB).

(72) Inventor: GOODFELLOW, Peter, Neville; 20 Lisburne Avenue, Belsize Park, London NW3 (GB).

(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS FOR IDENTIFYING A MUTATION IN A GENE OF INTEREST

(57) Abstract

The invention encompasses methods of identifying the presence of a mutation in a gene of interest in an organism which include identifying in a test nucleic acid sample from a mutated organism or a mutated cell from an organism a mutation in a gene of interest without the prior observation of a phenotypic alteration in the mutated organism or cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHODS FOR IDENTIFYING A MUTATION IN A GENE OF INTEREST

Field of the invention

The invention relates in general to methods for the identification and characterization of a mutation in a gene of interest.

Background to the invention

Detection of non-naturally occurring nucleotide sequence mutations has been approached by performing studies on cells in culture or on live animals based on alterations in phenotype. Tests on cells in culture using bacterial or animal cells or cell lines permits the rapid screening of a large number of cells for the appearance of an altered phenotype. The appearance of an altered phenotypic trait reflects the occurrence of a mutation in the test gene.

Previous attempts to identify genetic mutations have involved genetic mutation analysis based on phenotypic screening (Russell *et al.*, 1979, *PNAS* 76:5818; Russell *et al.*, 1982, *PNAS* 79:3589). That is, a phenotypic abnormality, such as alteration from wild-type (*eg.* coat colour in mice), is detected in the F1 offspring of a mutated animal, or in subsequent generations. Thus, Russell *et al.* assess mutation frequencies in a number of loci by identifying a mutant phenotype and correlating phenotype with a mutation at a corresponding locus. This is known as the "specific-locus method" of calculating the frequency of mutations in a given locus. However, observation of a mutant phenotype does not directly identify the gene which is mutated, although for phenotypes known to be the result of mutation of a particular gene, it may be inferred and subsequently tested. Phenotypes of interest can serve as a guide to study particular genes, using conventional mapping and positional cloning techniques to identify a gene or genes relating to the phenotype. This approach relies on the occurrence of a phenotype which is

used to score for a mutation, and the phenotype acts as a guide to the mutated gene.

Johnson *et al.* (1981, *PNAS* 78:3138) and Lewis *et al.* (1985, *PNAS* 82:5829) disclose a protein phenotype screen which detects electrophoretic mobility changes in proteins to test for induced genetic mutations. Protein extracts are isolated from a number of mutagenised animals, and specific proteins are assayed to look for abnormal electrophoretic migrations. This system identifies a change in phenotype in a protein in order to find a mutation in its corresponding gene.

The disadvantage of phenotypic screening for gene mutations is that the analysis of mutation distribution is always based on the window of observation that is permitted by the selective mutation system used, in which an alteration in a cell phenotype indicates that a mutation has occurred in a particular gene. The chief drawback of mutation assays involving phenotypic selection is that mutation analysis is confined to those genetic alterations which produce an altered gene product which is detectable via a phenotypic screen. Therefore, a phenotype must be matched with the mutation prior to detection or characterization of the mutant gene itself. For example, US Patent 5,347,075 discloses mutagenesis testing using a transgenic animal carrying a *lacZ* reporter test gene, wherein either cells containing the test gene or the animal itself is mutated, the "mutated" test gene is cloned in bacteria and then grown on X-gal indicator plates. Mutations in the reporter test gene are indicated phenotypically as white plaques rather than blue plaques.

Previous attempts to identify genetic mutations have also involved purely genotypic mutation analysis *in vitro*, or non-phenotypic selection of mutations (Palombo *et al.*, 1992, *Nucl. Acids Res.* 20:1349; Chiocca *et al.*, 1992, *PNAS* 78:3138). In these analyses, cells in culture are mutagenised, the DNA isolated, and tests are performed to detect mutations, for example, via changes in specific

restriction endonuclease sites (RFLP analysis). Although this procedure tests DNA directly for induced mutations, it has been adapted solely for mutagenesis of cultured cells.

US Patent 5,045,450 discloses a method of determining a mutation spectrum in a DNA sequence of interest that is present in a population of cells. The method includes detecting spontaneous mutations in a DNA sample wherein DNA is extracted from the tissue to be analyzed, hybridized to form duplexes with non-mutated DNA, and subject to denaturing gradient gel electrophoresis to detect mutations.

Mutation detection can be divided into two categories, each of which have different requirements: the detection of mutations in candidate disease genes; and the identification of mutations in known disease genes. The detection of mutations in candidate disease genes has been based on the mapping of a particular phenotype to a particular chromosome region and the examination of all genes mapping to this region for mutations in order to identify the gene responsible for the disease.

Animal models of disease have been produced in the prior art via phenotypic observation of a mutated animal. See, for example, Harding *et al.*, 1992, *PNAS* 89:2644, in which a mouse mutant with sarcosinemia was found by screening the progeny of ENU-mutagenised mice for aminoacidurias; and Bode *et al.*, 1988, *Genetics* 118:299, in which ENU-mutagenesis was used to screen for defects in phenylalanine metabolism by detecting elevated serum levels of phenylalanine. Mouse models of disease have also been produced via targeted mutagenesis involving targeting of a specific gene in ES cells, production of a mouse from the mutated ES cells, and ascertainment of phenotype. In such targeted mutations, the mutated gene is typically a "knockout" *ie.* in which a mutation is generated which inactivates the gene. For example, see Sadlack *et al.*, 1993, *Cell* 75:253, in which mice deficient for IL-2 were constructed; and Colledge *et al.*,

1995, *Nature Genetics* 10:445, in which mice were generated carrying a mutation in the cystic fibrosis gene.

One object of the invention is to provide mutational screening methods based on genomic and genetic techniques, rather than on phenotypic observation, to identify and characterize a mutation in a gene of interest.

Another object of the invention is to identify and characterize genes via mutagenesis in order to identify genes encoding products which may have therapeutic benefit.

Another object of the invention to provide methods for identifying the presence of mutations in a gene of interest which do not rely solely upon prior matching of a gene with a disease.

Another object of the invention is to provide methods for identifying the presence of mutations in a gene of interest which do not rely upon prior matching of a phenotypic mutation to a gene.

Another object of the invention is to allow the screening of heterozygotic organisms to identify those carrying a mutant copy of a gene of interest, even though the mutation may be recessive and therefore causes no phenotypic alteration.

There is a need in the art for a direct test for the presence of mutations in the genome of animals without using a phenotypic guide.

Summary of the invention

The invention is based on the discovery that the presence of a mutation in a gene of interest in an organism may be identified without first observing the mutated organism with respect to a phenotypic effect of the mutation. Accordingly, the invention provides a novel method of identifying the

presence of induced mutations in any particular gene in the genome of an organism without observing the phenotypic effects of the mutation prior to identifying the mutation.

The inventive methods are particularly advantageous in that they permit direct testing of nucleic acid, and are thus independent of screening for a potential phenotype resulting from mutations in the tested gene.

In one aspect, the invention encompasses a method of identifying a mutation in a gene of interest in an organism, comprising testing a nucleic acid sample from a mutated organism for a mutation in a gene of interest without the prior observation of a phenotypic alteration in the mutated organism (*ie.* an alteration resulting from the mutation).

Once the presence of a mutation in a gene of interest in an organism has been identified in this way, the mutation can be characterised and the phenotypic alterations resulting therefrom can be investigated.

The method may further comprise prior to the testing step, the step of mutagenising an organism so as to produce the mutated organism.

The invention also encompasses a method of identifying a mutation in a gene of interest in an organism, comprising the steps of (in the following order): mutagenising an organism to produce a mutated organism, testing a nucleic acid sample from a mutated organism for a mutation in a gene of interest without the prior observation of a phenotypic alteration in the mutated organism, and optionally subsequently observing the phenotype of an organism which has been identified as containing a mutation in the gene of interest. The phenotype in question is one resulting from the mutation.

In another aspect, the invention encompasses a method of identifying a mutation in a gene of interest in an organism,

comprising the steps of (in order): mutagenising a plurality of the same organism to produce mutated organisms; and testing a mixture of pooled nucleic acid samples containing a plurality of nucleic acid samples from a corresponding plurality of mutated organisms for the presence of a mutation in a gene of interest without the prior observation of a phenotypic alteration in the mutated organisms resulting from the mutation, and optionally observing the phenotype of the organism containing the mutated gene of interest.

As used herein, a "plurality" refers to a large number of organisms, *eg.* 100, 1000, 10000 and even up to 100000.

The mutated organism (*ie.* the organism whose nucleic acid sample is tested for the presence of a genetic mutation) may be the same as the organism which is mutagenised; alternatively, the mutated organism may be the offspring of the organism which is mutagenised, for example the F1 generation of the organism which is mutagenised, or the F2 or a subsequent generation. If, after a mutation is detected in an organism, it is desired to generate offspring from it, it is obviously preferable that the mutated organism be the offspring of a mutagenised organism. This ensures that the mutations carried throughout its somatic and germline cells correspond, such that any offspring from the mutated organism carry the same mutation as was detected.

Preferably, the testing step comprises hybridization of a nucleic acid probe to the sample or the mixture, the probe being unique to the gene of interest. Alternatively, the testing step comprises hybridization of a mixture of multiple nucleic acid probes to the sample or the mixture, the multiple nucleic acid probes differing in sequence from each other and being unique to the gene of interest.

As used herein, "nucleic acid" may refer to DNA, such as genomic DNA or cDNA, and also to RNA.

In another aspect, the invention encompasses a method of identifying a mutation in a gene of interest in an organism, comprising mutagenising the germline DNA of an organism; mating the mutagenised organism to produce F1 offspring; and testing a nucleic acid sample from an F1 offspring for the presence of a mutation in the gene of interest without the prior observation of a phenotypic alteration in the F1 offspring resulting from the mutation.

The germline DNA may be from a female or male organism, but is preferably from a male organism.

In another aspect, the invention encompasses a method of identifying a mutation in a gene of interest of an organism, the method comprising the steps of a) providing a mixture of pooled nucleic acid samples, each nucleic acid sample of the pool being from a mutated organism; b) providing a nucleic acid probe unique to a gene of interest; c) testing the mixture for the presence of a mutation in the gene of interest by hybridizing the probe to the mixture without the prior observation of a phenotypic alteration in each the mutated organism resulting from the mutation.

The method may further comprise the steps of d) detecting the presence of a mutation in the mixture; and e) testing each nucleic acid sample individually for a mutation in the gene of interest.

The method also may further comprise prior to step a), the step of mutagenising a plurality of organisms.

Thus the invention provides a method of identifying the presence of a mutation in a gene of interest in an organism, comprising mutagenising a plurality of the same organism to produce mutated organisms; and testing a mixture of pooled nucleic acid samples, said mixture comprising a plurality of nucleic acid samples from a corresponding plurality of mutated organisms for a mutation in a gene of interest

without the prior observation of a phenotypic alteration in said mutated organisms resulting from said mutation.

In any of the above-described inventive steps each mutated organism has been mutagenised such that about 1 mutation may occur as low as once in every 10000-1000 genes, but preferably occurs once in every 1000-500 genes.

In another aspect, the invention encompasses a method of identifying the presence of a mutation in two or more genes of interest in an organism, the method comprising the steps of a) providing a nucleic acid sample from a given tissue of a mutated organism, wherein the mutated organism contains about 1 mutation in 1000 genes; b) providing plural nucleic acid probes, each probe being unique to a given gene of interest; and c) testing the nucleic acid sample for the presence of a mutation in a gene of interest.

The method also may comprise prior to step a) the step of mutagenising an organism to produce the mutated organism.

Preferably, the mutagenising step in any of the above-described methods comprises inducing a genetic mutation into a gene of interest in an organism at an average frequency of 1/500, preferably 1/1000-1/10000 organisms.

Preferably, the mutation in any of the above-described methods is a single base pair mutation or a short insertion, deletion or substitution mutation, for example, in the range of about 1-10 base pairs.

One preferred method of mutagenesis according to the invention comprises exposing the organism to an alkylating agent, such as ethyl- or methyl-nitrosourea (ENU or MNU).

One preferred method of mutagenesis according to the invention involves mutating germline DNA of the organism.

It is preferred according to the invention that the probe or probes comprises a pair of unique PCR primers, and that the testing for the presence of a mutation in the gene of interest comprises amplification of a segment of the gene of interest and sequencing of the amplified segment.

One preferred method of testing for the presence of a mutation in the gene of interest is fSSCP analysis.

As used herein, the term "organism" refers to multicellular eukaryotes that undergo development from an embryonic stage to an adult stage. Accordingly, this includes vertebrates and invertebrates, which fall within the term "animal", as well as plants and fungi. The invention is useful with respect to animals, such as insects, nematodes, fish, such as zebrafish, or mammals, for example a rodent such as a mouse or a rat.

The invention also encompasses a method of identifying the presence of a mutation in a gene of interest in a tissue, comprising mutagenising an embryonic stem (ES) cell to produce a mutated ES cell; and testing a nucleic acid sample from the mutated ES cell for the presence of a mutation in the gene of interest without the prior observation of a phenotypic alteration in the mutated ES cell resulting from the mutation.

The invention also encompasses a method of identifying the presence of a mutation in a gene of interest in a tissue, comprising mutagenising plural ES cells to produce a plurality of mutated ES cells; and testing a nucleic acid sample from each mutated ES cell or a nucleic acid sample comprising nucleic acid from a plurality of mutated ES cells for the presence of a mutation in the gene of interest without the prior observation of a phenotypic alteration in the mutated ES cell resulting from the mutation.

It is preferred that the steps of these methods be performed in their stated order.

Preferably, in these methods, the testing step includes PCR amplification and fSSCP analysis using a pair of PCR primers from a region of the gene of interest. Preferably, the methods also include the steps of transferring the mutated ES cell to a developing embryo of the same organism species from which the ES cell is derived or of transferring the nucleus of the mutated ES cell to an egg of the same organism from which the chromosomes have been removed; and permitting the embryo to develop into a newborn.

Methods of the invention provide a mutagenised organism containing a mutation in a gene, the presence of which may be identified significantly more rapidly and at a lower cost than an analogous organism generated using for example transgenic technology.

Identification of an organism containing a mutant gene according to the invention permits the subsequent assessment of a phenotype resulting from the alteration of gene function, and provides a model organism to further disease diagnosis and drug development for both human and non-human diseases. A mutant gene identified according to the invention, or its wild-type counterpart, may encode a product which is useful as a therapeutic or as a target for a therapeutic, or for identifying a corresponding human therapeutic target.

The invention also is useful for identifying a sequence in a first organism, for example, a human, wherein very little, if any, sequence information is available, but where a disease or syndrome is known in a second organism of a different species, and it is desired to identify a corresponding gene or genes in the first organism.

The invention also therefore encompasses a method of identifying a gene of interest in a first organism, comprising in order the steps of: a) identifying in a second organism of a species different from the first organism a

mutation in a gene of interest in a nucleic acid sample from a mutated second organism without the prior observation of a phenotypic alteration in said second organism resulting from the mutation; b) correlating the mutation with a phenotypic alteration in said second organism resulting from the mutation; and c) utilising the gene of interest from the second organism, or sequences thereof, to identify the gene of interest in the first organism. After the gene has been identified, it may be isolated for further study.

Also according to the invention, a gene of interest can be obtained from a first organism after the inventive methods comprising mutagenesis have been applied to a second organism of a species different from that of the first organism. The second organism is then mated to produce offspring, which are then screened without prior observation of a phenotype for the presence of a mutation in the gene of interest. Once the presence of a mutation in a gene of interest has been identified, the mutation may then optionally be correlated with a phenotypic alteration. The gene of interest or sequences thereof from this second organism are then used to probe a nucleic acid sample from the first organism to obtain the gene of interest from that first organism.

In these embodiments of the invention, it is preferred that the first organism is a human and that the second organism is a mouse. It is also preferred that the testing comprises hybridization of a nucleic acid probe or mixture of multiple nucleic acid probes, wherein the probe or multiple probes are unique to the gene of interest and multiple probes differ in sequence from each other (for example, degenerate probes, or probes corresponding to different regions of a gene of interest or overlapping regions within the gene of interest, to a nucleic acid sample or mixture of pooled nucleic acid samples from the offspring of the mutagenised second organism).

The combination of organism mutagenesis and highly efficient

mutation detection according to the invention permits the analysis of a range of different mutations in single genes, and enables analysis of classes of genes, such as gene families and genes known or suspected to be commonly involved in a developmental process or disease. This includes candidate genes identified through positional cloning experiments.

The inventive screening methods confer significant advantages over prior art methods in that the inventive methods are significantly less expensive and significantly faster.

As used herein, "mutation" refers to an alteration in the nucleotide sequence of a given gene or regulatory sequence from the naturally occurring or normal nucleotide sequence. A mutation may be a single nucleotide alteration (deletion, insertion, substitution), or a deletion, insertion, or substitution of a number of nucleotides. The term "mutation" also includes chromosomal rearrangements.

"Induced mutation" refers to those mutations which are caused to occur by subjecting an organism, or cells of its germline, to a mutation-inducing condition, whether the inducing agent be a chemical or other mutagen or a gene mutation which induces mutations in the genome. For example, an "induced mutation" according to the invention may occur as a result of the use of a chemical mutagen or radiation mutagenesis in the laboratory. In addition, an "induced mutation" according to the invention may occur as a result of a mutation in a housekeeping gene of an organism which gives rise to additional mutations in the genome; for example, a mutation in a gene which encodes a DNA repair enzyme gives rise to numerous additional mutations in the genome of the organism. Induced mutations thus encompass non-naturally occurring mutations and do not encompass spontaneous mutations, which are defined by their exceedingly low frequency of occurrence (<1/100000).

"Phenotype" refers to the biological appearances, including chemical, structural, and behavioural attributes of an organism, such as an organism or tissue thereof, and excludes its genetic constitution. "Genotype" defines the genetic material that an organism inherits from its parents. The phenotype changes with time as the appearance of an organism changes, whereas the genotype remains relatively constant except for genetic changes known as mutations. Phenotypic information refers to both obvious changes in the visual appearance of an organism (*eg.* coat colour) and also to less obvious changes, such as in cellular growth of a tissue of an organism or a cultured cell line (*eg.* the adaptive ability to grow in the presence of a particular toxic chemical, or alterations in the electrophoretic mobility of a protein).

Where mutational screening is performed on tissue from an organism that has been subjected to induced mutagenesis "without phenotypic information" or "without regard to phenotype" of the mutated organism, a genetic, genotypic, or gene analysis (*ie.* all referring to analytical techniques based on nucleotide sequence or nucleic acid analysis) is performed prior to any optional observation of phenotype resulting from the induced genetic mutation. That is, where a nucleic acid sample is tested for the presence of a mutation in a gene of interest "without the prior observation of a phenotypic alteration" in the mutated organism, this means that, following mutagenesis of DNA, there is no testing, detection, or selection of an associated phenotype, *ie.* observations as to changes in subcellular (other than changes in DNA sequence or modification), cellular or organism behaviour, metabolism, *etc.* Samples to be screened need not have been selected on the basis of prior observation of phenotypic alterations in the organisms from which they are derived.

The term "gene" refers to a segment of DNA which may be transcribed into RNA, and which may contain an open reading frame and encode a protein, and also includes the DNA

regulatory elements which control expression of the transcribed region. Therefore, a mutation in a gene may occur within any region of the DNA which is transcribed into RNA, or outside of the open reading frame and within a region of DNA which regulates expression of the gene (*ie.* within a regulatory element). In diploid organisms, a gene is composed of two alleles. Preferably, a mutated diploid organism used in the inventive methods is heterozygous *ie.* it only carries a mutation in half of its chromosomes and thus only one allele is mutated.

The methods of the invention also may consist essentially of the described steps. Unless indicated otherwise, "consist essentially of" refers to a series of steps which include a step of testing or screening for the presence of a mutation in a gene of interest and which exclude a step in which observation of a mutant phenotype in an organism or tissue thereof resulting from the mutation is performed after mutagenesis of the organism or tissue but prior to testing for the presence of a mutation in the gene of interest.

The invention also provides a method for identifying those members of a mutated population which carry a mutation in a specific gene of interest, comprising the step of screening nucleic acid samples from a plurality of mutated organisms, without prior observation of phenotypic alterations in said population.

The invention also provides a method for detecting a member of a heterozygous mutated population which carries a mutation in a gene of interest, comprising the step of screening nucleic acid samples from a plurality of mutated organisms, wherein said plurality of mutated organisms are heterozygous for mutations. Preferably said mutation in said gene is recessive and is not manifested as a phenotype. A suitable population can be produced by mating a mutagenised organism with a non-mutagenised organism.

The invention further provides a method for selecting an individual from a plurality of mutated organisms, comprising the steps of: screening nucleic acid samples from the organisms; identifying a nucleic acid sample which contains a mutation in a gene of interest; and selecting the individual from which that sample was derived. The screening is carried out without organisms from which samples should be taken being selected on the basis of an altered phenotype.

The invention also provides a method for producing an organism carrying a mutation in a gene of interest, comprising the steps of: producing a plurality of mutated organisms; screening nucleic acid samples from said organisms in order to identify an individual carrying a mutation in a gene of interest; and breeding from said individual to produce offspring carrying said mutation. The plurality of mutated organisms are preferably themselves the offspring of an organism which was exposed to mutagenic conditions.

Further features and advantages of the invention will become more fully apparent in the following description of the embodiments and drawings thereof, and from the claims.

Brief description of the drawings

Figure 1 is a diagram of the *Sox3* gene showing the location of primers which are useful according to the invention for detection of mutations in the *Sox3* gene.

Description of the invention

The current challenge in biology is in understanding the function of the 50-100,000 unique human genes in normal development and the role any given gene may play in disease. The identification of genes is proceeding rapidly, but ascribing function to any particular gene is slow. For example, the use of expressed sequence tags (ESTs) has rapidly provided over 600,000 expressed sequences, very few

of which have been characterized. For some of these gene sequences, similarity to a gene family of known function may suggest a possible function, but determining a function for the gene is difficult and laborious. Another important advance has been the development of positional cloning to identify specific human genes involved in human disease. Genes identified by both sequence similarity and by positional cloning may potentially serve as therapeutic agents or targets for therapeutics if the gene function can be identified. In either case, it may be straightforward to deduce the sequence of the gene product, but deducing its mode of action is not straightforward. The biological role of proteins is best studied *in vivo* using animal models, as humans do not lend themselves well to these studies. A traditional route to ascribing function to a given gene has been to use phenotypic analysis of potential gene mutants to identify an aberrant phenotype and correlate the aberrant phenotype with a mutation in a gene or at a locus suspected to contain the gene.

The invention provides a first step in the analysis of gene function by permitting identification of a mutation in a gene of interest which may not have an ascribed function, but for which at least some nucleotide sequence information is known, *ie.* at least enough to provide a unique probe for the gene. The invention thus provides a much preferred alternative to traditional genetic analysis of gene function, wherein gene mutations are induced and their existence identified initially via their phenotypic effects.

Non-human mammalian models, for example, a rodent such as the mouse, are preferred model organisms for the study of mammalian gene function and are commonly used as model systems for human disease. This is due to the similarity of the human and non-human mammalian developmental and biochemical pathways (which is reflected in similarities in genes and genomic structure) and to the relative ease with which some mammals, such as mice, rats, hamsters, *etc.* can

be housed and bred. Investigation of the mammalian equivalents of human genes allows determination of the spatial and temporal pattern of gene expression during critical developmental time points and in specific tissues, contributing to the understanding of protein function. However, it is by studying abnormal forms of the gene *in vivo* that the scope of the gene function can be assessed and therapeutic treatments can be developed. The most powerful tool for determining the biological function of a gene is the use of targeted mutagenesis, a procedure which has been developed for different organisms, ranging from invertebrates to vertebrates. This technology has been used to modify specific genes.

1. Prior attempts to modify the mouse genome

Direct evidence for gene function has come from the observation of phenotypes conferred on individuals as the result of mutations in that gene. The phenotypic changes produced in the mutated organism can provide insight into the activity of a given gene. In some instances, gene modifications were made specifically to provide an animal model for human disease, *eg.* for cystic fibrosis. In other cases, genes have been modified (usually inactivated) as a means to determine their function. In some cases, the modification may produce a phenotypic change which can be related to the human disease. For example, the interleukin-2 gene produces a hormone which is thought to have a key role in the immune response of mammalian cells. The function of this gene was examined by deleting it through targeted mutagenesis in the mouse. Half of the mice died from immune system dysfunction, but all surviving mice develop an ulcerative colitis with striking clinical and histological similarity to inflammatory bowel disease in humans, providing evidence for a primary role of the immune system in ulcerative colitis and a new focus for the treatment of this disease in humans.

The invention provides advantages over the above-described

prior art in that it provides methods of gene analysis which ultimately lead to a better understanding of the full range of activity and function of a gene for which a function may already be ascribed, such as those described above. The invention permits the identification of numerous mutations in a gene of interest and the generation of, for example, mutant organism gene homologs of a human gene of interest without relying on an aberrant phenotype to identify the presence of a mutation in the gene.

2. The Inventive Methods

The invention is based on the discovery than any given gene of interest in an organism may be tested for the presence of a mutation in the gene without first observing the mutated organism with respect to a phenotypic effect of the mutation.

According to the invention, a mutation in a gene of interest is rapidly identified in a mutagenised organism via nucleic acid analysis. In a preferred embodiment of the invention, mutations are chemically induced in the germline of an organism. Breeding of the mutagenised males to normal females results in offspring with one set of genes which is normal (from the mother) and one set of genes which has been exposed to the mutagen (from the father). Therefore, any mutations are heterozygous (present in only one of two copies of each gene) in the first generation (F1) organisms. As an alternative to breeding, nuclear transfer may be used according to the methods described in WO97/07669 and Wilmut *et al.*, 1997, *Nature* **385**, 810-13. A gene of interest is tested for mutations in a sufficient number of organisms to identify those with mutations in the gene. Thus, it is advantageous according to the invention to generate mutations in the organism's DNA at high frequency, and to perform mutational analysis on large numbers of organisms.

3. The inventive screening methods are significantly more powerful than conventional transgenic techniques.

Each mammalian gene has natural variants (alleles) within a

population. Inheritance of unique combinations of alleles from each parent result in offspring with unique phenotypes. Some variants may cause, contribute or predispose to disease. Mutations in a gene which result in complete loss of function (a null allele) will often have a severe effect on phenotype. However the null allele is not the only disease causing modification. Other, more subtle changes to the genes may produce drastically different phenotypes, and different mutations in the same gene can cause different disease. For example, in humans, certain mutations of the *WT1* gene can result in the development of kidney tumours, while other *WT1* mutations result in the failure of male sexual development in chromosomal male individuals. Similarly, different mutations in the human *FGFR2* gene cause the three distinct diseases; Pfeiffer syndrome, Crouzon syndrome and Apert syndrome. It is more appropriate to consider genes as having a wide, if not continuous phenotypic potential depending on where the modification is found within the gene. Prior art methods of identifying a mutation in a gene of interest such as *WT1* or *FGFR2* in an organism require observation of a number of organisms for potentially vastly different phenotypes.

An advantage of the methods of the invention is that a typical screen of 10000 organisms is expected to identify 5-15 independent and different protein altering mutations for each gene tested. No phenotypic screening is necessary to identify the mutations. The screen may also be performed on successive new organisms at a rate of 2000 per month until the appropriate type or number of mutations is identified. The ability to study an "allelic series" of mutations in a particular gene is crucial to the understanding of the full range of disease phenotypes associated with the gene. The organisms carrying these mutations may also be interbred providing a vast array of possible combinations, any of which may give insight into the disease state in humans.

4. The invention provides a complementary technique to

positional cloning

Positional cloning relies on the ability to associate a region of the genome with a particular genetic disorder. Some gene research is based on the premise that disease genes can be mapped and cloned by identifying regions of the human genome that are associated with human disease phenotypes. This is done by collecting large numbers of phenotypically affected and control individuals and testing the entire genomes for markers which show linkage to the disease phenotype. The result of the analysis is the identification of a region of the genome which presumably contains genes contributing to the disease. This region is then searched for genes which become candidates based upon their position in the genome.

One of the major obstacles for positional cloning is in determining which of a number of candidate genes is the correct gene. The standard approach to assessing the candidate genes for their role in disease is to investigate the gene in individuals with the disease. Mutations in the gene that correlate with the presence of disease provide evidence for involvement of the gene in the disease. This analysis can be confounded by the presence of natural variation in the gene, or can be exceedingly complex where the disease is the result of complex interactions of genes such that any given individual will have mutation in a subset of all of the possible genes which can lead to the disease. Where more than one gene contributes to the disease phenotype, many of the patients will have a normal gene at that location and mutations elsewhere in the genome. In these cases the mutations may be more subtle variations that are difficult to detect rather than mutations that directly obviate gene function.

Methods of the invention have advantages which alleviate these difficulties. Genes identified as candidates through positional cloning can be rapidly tested according to the invention by first identifying organisms with mutations in

the candidate genes, and then subsequently testing these organisms for the disease phenotype. The invention therefore also provides methods for determining the function of a whole collection of candidate genes of unknown function. Where the mutated organism is an animal, mutant animals which are identified in this type of screen may then be interbred to produce animal models for complex polygenic diseases. The invention thus also provides animal models for diseases involving polygenic interactions.

Methods of the invention

It is contemplated according to the invention that large numbers of nucleic acid samples, *eg.* from a large number of different organisms, may be screened in a single procedure. For example, the invention contemplates high throughput screening of as many as 10000-100000 samples. Where a large number of organisms is screened, the limits of such screening is found in the limits of production of a large number of organisms. Therefore, it is preferred that the choice of organism for mutagenesis and mutation detection be based on convenience of handling a given organism and the number of organisms which may be subjected to mutagenesis and screening, or, where the F1 generation of organisms is tested, the number of organisms which may be produced, *eg.* by mating. It is also preferred that inbred organisms are used in mating, thereby ensuring that any difference in the DNA sequence of the offspring arises as a result of mutagenesis, and is not a natural polymorphism in the population.

Therefore, according to the invention, where a gene of interest is identified for mutation detection, for example a human gene of interest, an organism is chosen for mutagenesis, which is a good model for human disease, such as a mammal. Once the mammal is identified as being a good candidate for mutagenesis, a mutagenesis technique is selected according to the guidance provided below. After the animal is mutated, a body tissue is chosen for DNA extraction

and analysis of mutations in the gene of interest. Any one of a number of mutation detection techniques may be selected for identification of one or more mutations in the gene of interest.

Alternatively, where it is desired that fewer animals be subject to mutagenesis or where it is desired that more than one gene of interest be analyzed for mutations, the animals are mutagenised, DNA extracted from a selected tissue and subject to mutation detection using plural DNA probes for the gene of interest, each probe having a unique DNA sequence.

A DNA sample for analysis according to the invention may be prepared from any tissue or cell line, and preparative procedures are well-known in the art. Obtaining DNA samples from tails of mice, for instance, is straightforward and does not result in death. The DNA analysed may be in any form, such as genomic DNA or cDNA.

A probe useful according to the invention is a nucleic acid having a sequence that is unique to the gene of interest and which is preferably no longer than 30-40 nucleotides, and optimally less than 25 nucleotides, *eg.* 18 - 22 nucleotides, with a minimum of 10 nucleotides. The preparation and labelling of probes is well-known in the art.

Mutagenesis according to the invention

The invention encompasses mutagenesis of whole organisms or of a selected tissue of an organism including but not limited to, for example, mutagenesis of germline cells of an organism, such as sperm stem cells or ova, or mutagenesis of ES cells of an organism, or introduction of a mutant gene into an organism which results in an increased frequency of mutations in the genome. Following mutagenesis of an organism, the organism may be analyzed directly for mutations, or it may be mated and the offspring analyzed for a mutation in a gene of interest. Obviously it is preferred to analyse offspring in order to ensure that any mutation

which is detected can be predictably passed on to further generations. Alternatively, following DNA analysis of a specific tissue for a mutation in a gene of interest, such as mutated ES clones in culture, the cells are transferred to the developing embryo. Mutagens and mutagenesis techniques which are applicable to organisms or cell mutagenesis are described below.

1. Types of DNA mutations.

Mutations in DNA may be (a) large lesion mutations, such as chromosomal breaks, rearrangements, and large insertions or deletions (in the order of kilobases); (b) small lesion mutations, such as cytogenetically visible deletions within a chromosome; and (c) small alterations, such as point mutations, insertions and small deletions (in the order of several-tens of bases). Any type of mutation may be analyzed according to the invention, although mutations which do not result in complete deletion of the gene of interest are preferred.

The invention is most useful for analyzing the latter category of mutations, *ie.* point mutations, insertions and small deletions, and therefore it is preferred that the mutagenesis technique used to induce mutations according to the invention induce these types of mutations.

2. Selection of Mutagenesis Technique.

The selection of a mutagenesis technique useful according to the invention is dependent upon several factors. Some mutagens cause a wide spectrum of mutation types at a fixed condition(s). Some mutagens cause different types of mutations depending upon the mutagen dosage, mode of delivery, and the developmental stage at which the mutagen is administered to the organism. In addition, a mutagen may induce mutations at different frequencies depending upon the dosage regimen, mode of delivery, and the developmental stage of the organism or cell upon mutagen administration, all parameters of which are disclosed in the prior art for

different mutagens or mutagenesis techniques. In addition, a defect in a gene which in wild-type form prevents mutations from occurring or repairs mutations may result in the failure to repair DNA mutations and thus provide a mutagenised genome for analysis according to the invention. Finally, the mutation rate from tissue to tissue will vary.

A mutagen or method of inducing mutations is considered useful according to the invention which provides the highest number of mutations per genome which does not kill the mutated organism.

Therefore, the following guidelines are important for selection of a mutagenesis technique or a mutagen for use according to the invention. First, the number of potentially mutant organisms which are generated for screening must be technically feasible. Second, the technique used to screen the generated organisms for mutations in a given gene or genes must be technically feasible. Third, the type of mutation induced in a gene of interest must leave the gene intact in the genome to the extent that it is detectable as described herein, with small deletions/insertions/ substitutions, such as single base pair to several base pairs, being preferred. With these considerations in mind, it is possible to screen organisms which have been mutagenised at a high frequency or at a low frequency.

Those mutagens or mutagenesis techniques which result in mutations which occur within a gene or its regulatory elements are most useful according to the invention. Chemical mutagens which result in such mutations include alkylating agents which cause single nucleotide changes.

Therefore, according to the invention, mutations are induced in an organism at a high enough frequency such that the number of organisms needed to screen for a mutation in a gene of interest is not prohibitive. For example, it is particularly useful according to the invention to induce

mutations at a high frequency in order to decrease the number of organisms screened. ENU mutagenesis is particularly useful according to the invention. Extensive studies of ENU mutagenesis in mice have shown that a phenotypic trait induced by mutation of a single gene appears in ENU-treated mice at an average frequency of 1 per 1000 mice (Russell *et al.*, 1979, *PNAS* 76:5818; Favor *et al.*, 1983, *Mut. Res.* 110:367; and Favor, 1986, *Mut. Res.* 162:69). Thus, approximately 1000 mice are screened in order to detect a mutation in a particular gene. Although the ratio of 1/1000 has been calculated in the prior art based on phenotypic assays, it is the only way of assessing the relative mutational frequencies of mutagens or mutagenesis techniques useful according to the invention, as direct DNA analysis of the frequencies of mutations induced by a given mutagen or mutagenesis technique has not been performed. Because phenotypic mutation frequencies are based on DNA mutations which alter or destroy the function of a protein such that it causes a phenotypic change, the number of changes in the DNA of these mice in a given gene will be higher than 1/1000 due to "silent" mutations, *ie.* which do not result in a phenotypic change. The same type of mutation frequency is obtained using other chemical mutagens, such as MNU, PRC, and MMS. Additional mutagens which may be considered equally useful according to the invention are listed below.

Although the mouse is specifically embodied herein as a representative organism that is useful according to the invention for inducing mutations and screening for mutations in a gene of interest, the invention is not limited to the use of mice. For example, other rodents such as a rat or hamster also provide representative animal models. However, the invention is not limited to mutagenesis and mutational analysis of a rodent. Non-rodent animals are equally appropriate, for example, organisms such as insects, nematodes, or fish, such as the zebrafish or medaka fish.

The zebrafish has been used as a genetic system and

conditions for gamma-ray mutagenesis and screening are well established (Chakrabarti *et al.*, 1983, *Brachydonio Genetics* 103:109; Walker & Streisinger, 1983, *Genetics* 103:125). The advantages of zebrafish over the mouse for genetic analysis are its small size, the ability to house a large number of animals cheaply, and the large number (hundreds) of embryos produced from one female. The time from fertilization to gastrulation is only about 5 hours at 28°C; somites form between 10-20 hours; and by 24 hours postfertilisation, a recognizable animal with rudimentary eyes and brain has formed. Thus, the early development of this vertebrate takes only about as long as a phage plaque assay. Rossant *et al.*, 1992, *Genes Dev* 6:1, describe mutational strategies for mutagenesis of zebrafish, including ENU mutagenesis.

Briefly, a three-generation cross in which F2 females, heterozygous for a number of induced mutations, are backcrossed to their father and mated to their brothers to reveal homozygous mutant phenotypes. A locus-specific mutation frequency of 1/1000 gametes scored is achievable in zebrafish using ENU mutagenesis. Therefore, one would need to screen at least 3000 mutagenised gametes to approach saturation mutagenesis, and fewer than 2000 gametes, *ie.* on the order of about 1000 gametes to perform mutational analysis according to the invention. ENU and EMS mutagenesis has been used to induce mutations in isolated sperm from zebrafish (Halpern *et al.*, 1993, *Cell* 75:1; Solnica-Knezel *et al.*, 1994, *Genetics* 136:1401). The Medaka fish has also been subjected to ENU mutagenesis (Shiva *et al.*, 1991, *PNAS* 88:2545), and can also be used to practice the invention. Zebrafish have been used in large-scale mutagenesis to search for genes controlling development in vertebrates (Mullins *et al.*, 1994, *Curr Biol* 4:189).

In addition to mutagenised animals, lower organisms are useful according to the invention, such as mutagenised insects, *eg. Drosophila*. EMS mutagenesis has been performed extensively on *Drosophila melanogaster* (Ashburner, 1989,

Drosophila, A Laboratory Handbook; Grell *et al.*, 1981, *Drosophila Environ Mutagen* 3:381; Ondrej, 1971, *Drosophila melanogaster Mut. Res.* 12:159). Non-insect primitive organisms such as the round worm *Caenorhabditis elegans* may also be used according to the invention. EMS has been used to mutagenise *C. elegans* (Wood, 1988, The Nematode *C. elegans*, Cold Spring Harbor).

Non-mammalian organisms, such as fish, nematodes, and insects, are particularly useful according to the invention in identifying mutations in genes which are suspected to play a role in early development of the organism, *eg.* in embryonic development, such as pattern-forming genes, limb-forming genes, or organ-forming genes.

From the above description, it is evident that, in order to be useful according to the invention, mutations also may be induced in an organism at a lower frequency provided that a higher number of organisms or tissue samples from organisms are screened for a mutation in a gene of interest. The number of organisms tested is generally limited by the following: the number of mutant organisms that are generated, and the number of organisms that are screened. It may be possible to generate and screen a sufficient number of organisms to detect even an exceedingly low frequency of mutation, *eg.* 1 mutation/50,000 organisms - 1/75,000. Although screening for mutations which occur at a given frequency may be labour-intensive, a screening procedure must be employed which is feasible.

The invention therefore contemplates the use of any type of mutagenesis technique, including chemical or radiation mutagenesis, and mutagenesis techniques which are based on molecular biology, such as introduction into an organism of a gene encoding a defective DNA repair enzyme, retroviral insertion mutagenesis and promoter- and gene-trapping mutagenesis, as described below.

The invention is particularly useful where the mutagenesis results in germline mutations, *ie.* which are passed onto offspring which are tested for mutations, and therefore relates to mutations which are induced in the germline of a parent organism.

In a preferred aspect of the invention, a mutagenesis technique is employed which confers a mutation rate in the range of 1 mutation per 500 genes - 1 mutation per 10000 genes, or 1 mutation per gene per 100 organisms - 1 mutation per gene per 10000 organisms, optimally at least 1 mutation per 1000 genes, or 1 mutation per gene per 1000 organisms. It is desired according to the invention that the mutation frequency possess an upper limit that is below the frequency of inducing a dominant lethal mutation in every organism.

A) Chemical Mutagenesis and Mutagens.
Chemical mutagens are classifiable by chemical properties, *eg.* alkylating agents, cross-linking agents, *etc.* Any suitable chemical mutagen can be used to practice the invention.

ENU, MNU, procarbazine hydrochloride and chlorambucil are particularly useful for mutagenesis of male germ cells. Other useful mutagens include: cyclophosphamide, ethyl and methyl methanesulfonate (EMS & MMS), diethyl sulphate, acrylamide monomer, triethylene melamin (TEM), melphalan, nitrogen mustard, vincristine, dimethylnitrosamine, N-methyl-N'-nitro-nitrosoguanidine (MNNG), 7,12-dimethylbenz(a)anthracene (DMBA), ethylene oxide, hexamethylphosphoramide, bisulfan.

A comparison of specific-locus mutation rates induced by various chemicals can be found in Table 2 of Russell *et al.* (1989) *PNAS* 86:3704-3708.

ENU mutagenesis in particular
One particularly useful mutagen according to the invention is the chemical mutagen ethylnitrosourea (ENU). ENU may be

used to induce genomic mutations in any organism, including lower organisms such as insects and worms, as well as higher organisms such as vertebrates. Mutagenesis and DNA mutation screen also may be applied to other organisms which are used as model systems for human disease. Rats are a good candidate for practical reasons, *ie.* since mouse-based animal facilities are also suitable for rats. The inventive methods are easily applicable to the rat and can produce and identify mutations in specific rat genes.

Previous ENU mutagenesis experiments in mice used in excess of 500000 animals. The genes involved were assayed indirectly by observation of phenotypic changes in the mice. ENU is believed to produce mutations at random throughout the genome, and the frequency of mutations, determined for numerous genes, is in the range of 0.5-1.5 phenotypically mutant mice per 1000 mice for any given gene, irrespective of the gene screened. In the past, the presence of mutations could only be inferred on the basis of a phenotype in the mutated mice. Most of these mutations did not produce an obvious phenotypic change in the heterozygous state and required additional breeding to make the mutations homozygous (F2 and F3 generations) to observe the effect of the mutation. Mutagenesis and mutant screening according to the invention do not require a previously-determined mutant phenotype, as nucleic acid is analyzed directly for the presence of a mutation in the gene of interest. In 1000 mice, 0.5-1.5 mutations in any gene may be detected. By screening 10000 mice, it is thus possible to identify 5-15 mice, each carrying heterozygous mutations in a target gene. Any number of genes can be screened in these same 10000 mice.

Each F1 mutagenised animal, in addition to carrying the mutation being studied, has unknown induced "background" mutations. For ENU mutagenesis, the mutation frequency can be 1 phenotype-inducing mutation/gene/1000 mice, which means that a group of 1000 mice will collectively have every gene mutated such that it will produce a phenotype (in most cases

when the mutation is homozygous). Estimates of the gene number in mice range from 50000 to 100000, so for all of these genes to be mutated (collectively) in 1000 mice, each F1 mouse will have 50-100 mutations. These unselected mutations may cause a phenotype which could be mis-assigned to the gene being studied. The background mutations are easily removed using standard breeding techniques. Unselected background mutations are randomly lost with each generation of breeding to a non-mutagenised mouse. If an F1 mouse contains 100 mutations (one of which is selected *ie*. the one which has been detected and is being studied), breeding to a non-mutagenised mouse will yield F2 mice, each of which has 50 mutations. One of these is selected for, the other 49 are retained at random from the original pool of 99 background mutations in the F1 animal. Continued outcrossing to non-mutagenised animals for 7 or more generations will remove all of the unselected mutations, resulting in a mouse carrying only the selected ENU induced mutation, in the absence of ENU-induced background mutations.

Using ENU mutagenesis on mice, it is expected that the gene of interest will be mutated to produce a phenotype once in 1000 mice. If a given animal genome contains, for example, 100000 genes, then each ENU mutated animal will contain in its ENU mutated genome one protein-altering mutation in one allele of every 1000 genes.

B) Radiation Mutagenesis

In general, X-rays, gamma rays, neutrons, *etc.*, cause DNA breakage. Cellular repair mechanisms of DNA breaks may result in regions of DNA which contain large lesions, including rearrangements and deletions. Although analysis of other types of mutations are preferred according to the invention, analysis of radiation induced mutations, which tend to be larger in that they encompass more bases, are also encompassed by the invention. UV light-induced mutations, however, are largely single nucleotide alterations. Because UV light does not penetrate an animal, however, it is

generally used for inducing mutations in cells in culture (*eg.* ES cells) or on exposed tissues *eg.* eyes or skin.

In addition to chemical or radiation induced mutations, mutations may be induced in an animal using insertional mutagenesis techniques, as follows.

C) Retroviral Insertion Mutagenesis

Retroviruses can be used to cause insertional mutations, and retroviral insertions are usually simple and cause little or no alteration in surrounding host DNA. Retroviral vectors are easy to use, infect a wide variety of cell types, including ES cells, are stable through multiple generations, and do not cause rearrangements of the host genome when integrated. The mutation frequency from retroviral insertion is estimated at about 1 mutation/1.5x10^6 cells (Keuhn *et al.*, 1987, *Nature* 326:295). (For retrovirally induced mutations in the mouse, see Harbers *et al.*, 1984, *PNAS* 3:162; Soriano *et al.*, 1987, *Genes Dev* 1:366; and Gridley *et al.*, 1987, *TIG* 109:235). A detailed description of retroviral insertion mutagenesis can be found in *Methods in Enzymology*, vol. 225, 1990.

D) Promoter- or Gene-Trapping Mutagenesis

Entrapment vectors, first described in bacteria (Casadaban and Cohen, 1979, *PNAS* 76:4530; Casadaban *et al.*, 1980, *J Bacteriol.* 143:971) permit selection of insertional events that lie within coding sequences. Entrapment vectors can be introduced into pluripotent ES cells in culture and then passed into the germline via chimeras (Gossler *et al.*, 1989, *Science* 244:463; Skarnes, 1990, *Bio/technology* 8:827). Promoter or gene trap vectors often contain a reporter gene (*eg.* lacZ) lacking its own promoter and/or splice acceptor sequence upstream. If the vector lands in a gene and is spliced into the gene product, then the reporter gene is expressed. Enhancer traps have a minimal promoter which requires an enhancer to function, and contains a reporter gene. If the vector inserts near an enhancer, then the reporter gene is expressed.

The vector may be introduced into ES cells by electroporation or using a retrovirus. Activation of the reporter gene can only occur when the vector is within an active host gene and requires generation of a fusion transcript with the host gene. The reporter gene activity then provides an easy assay for integrations in expressed genes. These DNA integrations are highly mutagenic because they interrupt the endogenous coding sequence. It is estimated that the frequency of obtaining a mutation in some gene of any in the genome using a promoter or gene trap is about 45%.

E) Deficiency of a DNA Repair Enzyme

The invention also encompasses mutagenesis as a result of a deficiency in a DNA repair enzyme. The presence of a mutant DNA repair enzyme is expected to generate a high frequency of mutations (*ie.* about 1 mutation/Genres - 1 mutation /10,000 genes). DNA repair enzymes include but are not limited to topoisomerases, helicases, and recombinases. Examples cf genes encoding such enzymes include *MutH, MutS, MutL,* and *MutU*, and homologs thereof, including mammalian homologs such as *MSH 1-6, PMS 1-2, MLH 1, GTBP,* and *ERCC-1.*

McWhir *et al.*, 1993, *Nature Genetics* 5:217 describe a mouse containing a defective DNA repair enzyme resulting from a mutation in the DNA repair gene *ERCC-1.* Homozygous mutants died before weaning, whereas heterozygous mutants survived and were available for mating. It is contemplated according to the invention that a mammalian organism heterozygous for a mutant gene encoding a DNA repair enzyme may be used according to the invention.

Where the organism is not an animal, mutagenesis and breeding procedures may be adapted as necessary. For instance, to produce a mutant population of plants it may be desired to mutagenise pollen or embryos, which can then be used to produce a suitable plurality of mutated organisms (*eg.* Meinke 1991, *Dev Genet* 12:382). The totipotency of plant cells also facilitates the generation of further organisms carrying a

mutation of interest, both heterozygotes and homozygotes.

Mutation detection according to the invention

Mutation detection analysis is preferably performed on a number of DNA samples simultaneously in order to increase the efficiency of identifying as many mutations as possible in a given gene of interest, or as many mutations as possible in a given organism. It is contemplated according to the invention that three general approaches to mutation screening are particularly useful. First, a single gene is examined for mutations using a nucleic acid probe unique to that gene, and a number of mutant organisms are screened in order to provide mutation detection in that gene. Second, a single gene is examined for mutations using a mixture of unique nucleic acid probes (a multiplex probe) for that gene, and a number of mutant organisms are screened to provide mutation detection in the gene examined. Third, several genes of interest (*eg.* 2 to 3) are examined for mutations using a mixture of several probes, each of which is unique for a given gene.

Provided below is a description of a particularly useful combination of mutagenesis and mutation detection according to the invention. This combination involves ENU mutagenesis of male mice, mating to allow production of F1 offspring, and mutation detection using SSCP screening. Other mutation detection techniques which are useful according to the invention are also disclosed below.

<u>Single Strand Conformation Polymorphism (SSCP) Screening and Fluorescent SSCP Screening</u>

One approach to detecting DNA mutations in a mutagenised organism is single strand conformation polymorphism (SSCP) (Orita *et al.*, 1989, *PNAS* **86**:2766; Glavac *et al.*, 1993, *Hum. Mut.* 2:404; Sheffield *et al.*, 1993, *Genomics* 16:325), which is based on the principle that single-stranded DNA molecules take on specific sequence-based secondary structures (conformers) under nondenaturing conditions. This technique has proven useful for detection of multiple mutations and

polymorphisms. SSCP sensitivity varies with the size of the DNA fragment being analyzed and the optimum size fragment for sensitive detection by SSCP is approximately 150-300bp.

fSSCP Analysis

High throughput screening of a large number of samples is advantageously achieved by pooling and multiplexing of DNA samples in fluorescent SSCP (fSSCP) assays (Makino *et al.*, 1992, *PCR Methods Appl.* 2:10; Ellison *et al.*, 1993, *Biotechniques* 15:684). PCR products can be visualized and analyzed using an ABI fluorescent DNA sequencing machine, with different coloured fluorochromes for different primer pairs.

Although SSCP and fSSCP techniques are preferred according to the invention, any DNA mutation detection system can be employed to test for mutations (*eg.* Eng & Vijg (1997) *Nature Biotech* 15:422-426, especially 424-425). Further suitable DNA detection techniques, with example references, are:

Denaturing Gradient Gel Electrophoresis (Fischer & Lerman 1987, *Methods Enzymol* 155:482-501; van Orsouw *et al.* 1996, *Hum Mol Genet* 5:755-761; US patent 5,190,856)

Cleavage of Mismatches (Marshal *et al.* 1995 *Nature Genet* 9:177-183; Youil *et al.* 1995 *PNAS* 92:87-91; Cotton *et al.* 1988, *PNAS* 85:4397-4401)

CDCE analysis (Khrapko *et al.*, 1994, *Nucleic Acids Res.* 22:3:364)

RNase Cleavage (Winter *et al.* 1985 *PNAS* 82:7575-7579)

Heteroduplex Analysis (Nagamine *et al.* 1989 *Am J Hum Genet* 45:337-339; Keen *et al.* 1991 *TIG* 7:5)

Mismatch Repair Detection (Faham *et al.*, 1995, *Genome Research* 5:474)

Mismatch Recognition by DNA Repair Enzymes *eg.* MutS

Sequencing by Hybridization (Chee *et al.* 1996, *Science* 274:610-614)

Dot-blots and **Reverse dot-blots** (Saiki *et al.* 1986 *Nature* 324_163-6; Cai *et al.* 1994 *Hum Mut* 3:59-63)

Allele-Specific PCR or amplification refractory mutation

system (Newton *et al.* 1989 *Nucl Acid Res* 17:2503-16)
Primer-Introduced Restriction Analysis (Shibuta *et al.* 1994 *J Med Genet* 31:576-9; Jacobson 1992 *Am J Hum Genet* 50:195)
Oligonucleotide Ligation (Landegren 1993 *BioEssays* 15:761)
Direct DNA Sequencing (Gibbs *et al.* 1989 *PNAS* 86:1919-23)
Protein truncation (Roest *et al.* 1993 *Hum Mol Genet* 2:1719)
Mini-Sequencing or single nucleotide primer extension (Syvänen *et al.* 1990 *Genomics* 8:684-692; Ihalainen *et al.* 1994 *Biotechniques* 16:938-943)
5' Nuclease Assay (Livak *et al.* 1995 *Nature Genet* 9:341-2).
Representational Difference Analysis (Lisitsyn *et al.*, 1993, *Science* 259:946; and *Nature Genet* 6:57).
See also "Finding Mutations: The Basics" by JR Hawkins (IRL Press, ISBN 0 19 963611 7).

Examples

EXAMPLE 1

A mutant phenotype induced by mutation of a single gene appears in ENU-treated mice at an average frequency of 1 per 1000 mice. By performing DNA mutation analysis of a gene in thousands of ENU treated mice, it is predicted that several independent mutations will be found which have an effect on the function of the gene product. The mutations may be detected directly in genomic DNA or cDNA.

ENU mutagenesis of an animal may be performed as follows.

350 ten to twelve week old male C3H mice (G0) (GSF Forschungs Zentium Inst. For Mammalian Genetics, Oberschleissheim, Germany) are injected intraperitoneally with ENU (Serva, Heidelberg, Germany) in 1 ml or less of 55 mM phosphate buffer pH 6.0. Single doses are administered weekly for a total of 3 doses at 100 mg/kg animal weight. This treatment regime maximizes the mutation frequency without severely impairing fertility. Following a brief period of sterility, (8-12 weeks), these mice are permanently mated to C3H females. Mutagenised F1 offspring are produced at a rate of

2000 per month, based on a litter size of 4-6 animals. The G0 males are allowed to produce 50-75 F1 animals. Many more mice than this can be generated by a single G0 male, but going beyond this range dramatically increases the chance of producing offspring with recurrent identical mutations.

A tail clipping is taken from each mouse at weaning and used to prepare high molecular weight genomic DNA for genomic mutation analysis. From a 10 day old mouse, approximately one centimetre of the tail is removed and placed in 500 μl TB buffer (50 mM Tris-HCl, pH 8, 100 mM NaCl, 1% SDS, 600 μg/ml proteinase K) and incubated overnight at 55°C. The sample is then extracted with 500μl 1:1 (w/w) phenol/chloroform and precipitated with two volumes of ethanol. DNA pellets are resuspended in 500 μl water.

Following identification of a F1 mouse which contains a mutation in the gene under study, as described in Example 2, the F1 mouse can be examined for any phenotype associated with the mutation. The F1 mouse was generated by mating a mutagenised male mouse with a non-mutagenised female, so it is heterozygous for the mutation. Only dominant mutations will cause a phenotype in the F1 animal, but few mutations are dominant, and so there will usually be no associated phenotype in the F1 animal. Most mutations are recessive, causing a phenotype only when present in both alleles, and to study phenotypes caused by recessive mutations it is necessary to breed the mouse to homozygous animals. This can be achieved by crossing the heterozygote with a normal mouse to generate F2 offspring. Half of the F2 offspring will be heterozygotes, and half will not carry the mutation in question; these can be distinguished by screening the F2 generation for said mutation. Those F2 mice carrying the mutation are bred together and quarter of the resultant offspring will be homozygous for the mutation. These animals can be tested for a recessive phenotype caused by the mutation being studied.

EXAMPLE 2

Mutation detection analysis is performed as follows. For a comprehensive SSCP screen of a target gene in genomic DNA, the DNA sequence must be known together with the intron/exon structure in the mouse genome or, alternatively, where intron regions are not known, primers are designed for targeting exon regions only, *eg.* using cDNA sequence information, and are tested on genomic DNA. It is contemplated according to the inventive methods to screen genomic DNA from an organism, as this is the most cost effective procedure. Alternatively, where the structure and organization of the intron/exon region is not known, and genomic DNA screening is not preferred cDNA may be prepared from each organism and screened instead of genomic DNA.

For maximum mutation detection sensitivity by SSCP the target gene is PCR amplified in fragments of 150-300 base pairs. It is estimated that the average screening will cover 2,000 base pairs and will require up to 15 PCR amplifications. PCR primer pairs containing one fluorescently labelled and one un-labelled primer, or both primers labelled, are designed to amplify the entire coding region including intron/exon boundaries, with some overlap between adjacent amplicons to ensure screening of the primer regions. Each primer pair is tested and conditions optimized for their ability to amplify a unique fragment of the appropriate size (1-2 weeks for 15 primer pairs). 2000 genomic DNA samples may be prepared each month. These samples, together with DNA from 8000 F1 mice generated in the previous 4 months, are screened for the presence of mutations. All of the mice from which DNA is screened are kept alive for further (phenotypic) analysis of the mice containing mutations in the gene being screened. The DNA from 5 mutagenised mice is pooled and all five DNA samples concomitantly amplified in a single PCR reaction. Approximately 2,000 PCR reactions are required to screen 10,000 mouse DNA samples with each PCR primer pair (one amplimer).

The PCR amplifications will be performed in machines with multiple, independent thermocycling blocks each having 192 reaction capacity (1 day/amplimer). The amplification products will be assayed using a modified ABI 377 automated sequencing machine. Detection of mutations by SSCP is optimal at temperatures of 4-15°C and the current 377 automated sequencer can be modified by ABI technicians to enable the gels to be cooled.

At present there are 4 different fluorescent dye labels which are compatible (at least one additional dye currently under development). and can be distinguished by its unique fluorescence when run in a single lane.

Therefore, allowing one fluorescent dye for use as a size standard, three different amplimers may be run concurrently when using 4 different fluorochromes. In a gel, 15 samples can be run per electrophoretic run in each of 36 lanes per gel (of the ABI 377 machine), which is equal to 540 samples per gel. If a gene is analyzed in 10 amplimers, then 100,000 samples are tested (10 amplimers per gene x 10,000 mice). To analyze the 100,000 samples (one gene in 10,000 mice), 185 gels are used (100,000/540). Each ABI machine can run approximately 2 gels per day. Therefore, where 4 machines are employed, 185 gels are run in 23 days. Similarly, if a gene is analyzed using 15 amplimers, then the same type of calculation results in 277 gels; using 4 machines, the results are available in 35 days. This type of numerical analysis of feasibility of a given mutagenesis/mutation detection combination permits a determination of the technical feasibility of employing that combination of techniques. It is possible to use any mutation inducing agent, even one which produces as low as 1 mutation per genome, in a given combination, provided a mutation detection technique is available, *eg.* SSCP, and the capability of testing a large number of samples of a given gene or genes of interest.

If desired, the region of DNA containing the mutation identified as described herein may be cloned or, alternatively, the gene containing the PCR-amplified region, and thus the mutation, also may be cloned, using conventional cloning techniques.

EXAMPLE 3

It is contemplated according to the invention that a plurality of DNA samples from different organisms may be pooled and screened in a single mutation detection assay. For example, where the mutation detection technique is fluorescence SSCP, the number of samples which may be combined is determined as follows.

For high throughput screening, DNA samples from different organisms may be combined (pooled) prior to amplification of the gene of interest. Using fluorescence SSCP, for example, a mutated allele appears as an aberrantly migrating peak of fluorescence. In pooled samples from, for example, 5 organisms, if 1 allele of 1 organism is mutated then 1/10 of the signal (2 alleles per organism, or 10 alleles per pool of 5 organisms) is present in the shifted peak and 9 times that amount of signal is present in the normal (unmutated) migration peak. The sensitivity of fSSCP permits visualization of a peak corresponding to a shift in mobility (corresponding to a mutated allele) which is 1/10 the signal of the normal unmutated alleles.

According to the invention, it is contemplated that as few as 5 DNA samples from different organisms may be pooled, or a larger number of DNA samples from different organisms may be pooled, *eg.* 10 samples, 15 samples, 20 samples, 50 samples, etc. The limit of the number of different organism DNA samples which may be pooled is determined by the limits of detection of a given single sample above background in the detection technique used. For example, for fSSCP, data is produced as a peak of units of fluorescence. A peak of about 6000 units is near the practical maximum level of detection.

A peak of 50 units is easily detected above background fluorescence levels. Therefore, greater than a 100-fold range of minimum - maximum detection levels is permitted using fSSCP between peak heights. If amplification of 1 allele produces a value of 50 units of fluorescence, then a pool of DNA samples from 60 different organisms (*ie.* 120 alleles), where a DNA sample from a single organism contains a mutation in 1 allele, then the mutated allele produces a value of 50 fluorescence units and the combined 119 normal (unmutated) alleles produces a value of 5650 fluorescence units.

EXAMPLE 4

Sox-3 (both mouse *Sox-3* and human homologue *SOX3*) is a member of the *Sox* gene family, a family of about 20 genes which all contain approx. 240 bp DNA sequence corresponding to a 80 amino acid segment (which is called an "HMG box") that contains about 60% or greater amino acid similarity to the *SRY* gene. Outside the HMG box, the genes are very different. The HMG box is a DNA binding domain and the *SOX* genes which have been studied bind to DNA via this region of the protein and are likely modulators of gene expression, such as transcription factors. *Sox-3* is expressed in the developing central nervous system. The complete sequence is known (see Collignon *et al.*, 1996, *Development* 122:509 for *Sox3* sequence, and Stevanovic *et al.*, 1993, *Hum. Mol. Genet.* 2:2013 for *SOX3*) and has an open reading frame of 1125 base pairs. Fig. 1 is a diagram of the *Sox3* gene showing the location of primers which are useful according to the invention for detection of mutations. The 1150-bp ORF of the gene is represented by an open box, with flanking non-coding DNA represented by solid bars. Lines labelled A-I indicate fragments (amplimers) amplified by PCR using the primer pairs listed below. PCR primer pairs useful to generate amplimers in *Sox3* to detect gene mutations have the following sequences (F = forward; R = reverse):

A-F	GCACCTCCTTCCCGCCCC	(SEQ ID NO:1)
A-R	TTCGCGCCCCCGCTGCCC	(SEQ ID NO:2)

B-F	GAGTGGCGCGAACCCAGC	(SEQ ID NO:3)
B-R	TTGGGGTTCTCCAGGGCC	(SEQ ID NO:4)
C-F	GAACGCGTTCATGGTGTG	(SEQ ID NO:5)
C-R	TACTCCTTCATGTGCACC	(SEQ ID NO:6)
D-F	GGAGAAGCGGCCGTTCAT	(SEQ ID NO:7)
D-R	GCGGCGGCGACGGCGGCG	(SEQ ID NO:8)
E-F	CTCGCTGCCCGGCGGCCT	(SEQ ID NO:9)
E-R	GGCTGCGCGTAGCCCAGC	(SEQ ID NO:10)
F-F	GAACGGCTGGGCCAACGG	(SEQ ID NO:11)
F-R	GCGGCGTTCATGTAGCTC	(SEQ ID NO:12)
G-F	GGCCGGCCTGCAGTACAG	(SEQ ID NO:13)
G-R	GCGGCGGCGGCGGCGGCC	(SEQ ID NO:14)
H-F	CGCCGCCGCCTACGGGCA	(SEQ ID NO:15)
H-R	TACATGCTGATCATGTCG	(SEQ ID NO:16)
I-F	CATCCGTTCGCACTCGCA	(SEQ ID NO:17)
I-R	TCAGATGTGGGTCAGCGG	(SEQ ID NO:18)

The primers are located throughout the gene and are unique to *Sox3*.

Mutations in the *Sox3* gene are generated, identified, and analyzed according to the invention as follows. Mice are mutagenised using ENU mutagenesis, as described above. The mutagenised mice are tested by PCR with each primer set and fSSCP analysis, which identifies several mice carrying mutations in *Sox3*.

These mice may be examined and bred to observe their phenotype. Prior to ENU mutagenesis of mice, it was known that in human *SOX3*, there is an individual who has a deletion which removes (minimally) SOX3 and the factor IX (blood clotting) gene. This individual has mental retardation and haemophilia and *SOX3* may be linked to X-linked mental retardation. It is also known that *Sox3* mRNA in mice is abundant in the central nervous system. Therefore, in the

several ENU mutagenised mice for which a mutation in the *Sox3* gene is identified, any displaying neural defects provide a mouse model for human mutant *SOX3* mutations.

EXAMPLE 5

It is contemplated according to the invention that cells in culture are mutagenised prior to producing an organism and analysed for a mutation in a gene of interest *eg. Sox3*. One such example includes mutagenesis of ES cells, as follows.

ES cells are prepared from mice (as described in Gossler *et al.*, 1989, *Science* 244:463, and Skarnes, 1990, *Bio/technology* 8:827) and are mutagenised using retroviral mutagenesis, or gene- or promoter-trapping mutagenesis, as described above. If desired, however, any other mutagenesis method may be used *eg.* ENU or UV mutagenesis. 10000-1000000 mutagenised ES clones are screened by fSSCP analysis using one or more primer pairs, as described in Example 4, for a mutation in the *Sox3* gene, or by any other mutagenesis screening technique described herein. A mutagenised ES clone in which a *Sox3* mutation is identified is then introduced by micro-injection into a mouse embryo at the blastocyst stage, and the mouse is produced via normal gestation and birth. A mouse thus produced from a *Sox3* mutant ES clone will itself contain the Sox3 mutation, and thus be available for further phenotypic or biochemical studies.

As an alternative, the mutagenised ES clone of interest can be used according to the methods of Campbell *et al.* (WO97/07669), wherein these cells serve as nuclear donors in order to produce a cohort of genetically-identical, cloned mice. It is also contemplated that screening for the presence of mutations in the *Sox3* gene occurs after numerous independent ES cell clones are used to generate cloned mice.

EXAMPLE 6

Sox-2 (both mouse *Sox-2* and human homolog *SOX2*) is also a member of the Sox gene family, involved in transcriptional

regulation of the *FGF-4* gene. FGF-4 codes for a signalling protein whose expression is essential for postimplantation mouse development and, at later embryonic stages, for limb patterning and growth. *FGF-4* gene is expressed in the blastocyst inner cell mass and later in distinct embryonic tissues, but is transcriptionally silent in the adult. The mouse *Sox2* nucleotide sequence (see Yuan *et al.*, 199, *Genes Dev* 9:2635) has an open reading frame of 956 base pairs. Human *SOX2* cDNA is 1085bp long and contains a 317 amino acids ORF (Stevanovic *et al.*, 1994, *Mammalian Genome* 5:640.) and displays a high degree of similarity with mouse *Sox2*. PCR primer pairs useful to generate amplimers in the Sox2 gene to detect gene mutations have the following sequences (F = forward; R = reverse):

A-F	CCACAGTCCCGGCCGGGC	(SEQ ID NO:19)
A-R	GGGCTGTTCTTCTGGTTG	(SEQ ID NO:20)
B-F	CGGCGGCGGAGGAGGCAA	(SEQ ID NO:21)
B-R	TCCGCGCCCAGGCGCTTG	(SEQ ID NO:22)
C-F	GATGGCCCAGGAGAACCC	(SEQ ID NO:23)
C-R	GTCTTGGTTTTCCGCCGC	(SEQ ID NO:24)
D-F	CGCTCTGCACATGAAGGA	(SEQ ID NO:25)
D-R	CTGTCCATGCGCTGGTTC	(SEQ ID NO:26)
E-F	GGCGAGCGGGGTTGGGGT	(SEQ ID NO:27)
E-R	CGGTGCATCGGTTGCATC	(SEQ ID NO:28)
F-F	CTACCCGCAGCACCCGGG	(SEQ ID NO:29)
F-R	GAGCCCAGCGCCATACCG	(SEQ ID NO:30)
G-F	CTCGCCCACCTACAGCAT	(SEQ ID NO:31)
G-R	GGGAGGTACATGCTGATC	(SEQ ID NO:32)
H-F	CCACTCCAGGGCGCCCTG	(SEQ ID NO:33)
H-R	CCTCACATGTGCGACAGG	(SEQ ID NO:34)
I-F	GTGCGGCCCGGTGCCCGG	(SEQ ID NO:35)
I-R	AACCACCAAAAAAAGGAA	(SEQ ID NO:36)

Mice carrying *Sox2* mutations may be examined and bred to observe their phenotype and thus to gain more information on *Sox2* and *SOX2* gene function. Prior to mutagenesis of mice, very little is known about human *SOX2* homolog other than its expression in fetal brain tissue. More is known with respect to *Sox2* function as it relates to transcriptional regulation of *FGF-4*. Therefore, in those mutagenised mice for which a mutation in *Sox2* is identified, the phenotype of the *Sox2* mutant mice are observed for, for example, postembryonic fetal development and limb formation, thus providing a mouse model for a mutant human *SOX-2* gene.

Alternatively, instead of mutagenising a mouse and analysing its DNA, or that of an F1 offspring, for Sox2 mutations, ES cells from a mouse may be mutated and analyzed using the same *Sox2* primers, as described in Example 5 for *Sox3*.

EXAMPLE 7

A *Drosophila* gene is provided which appears to function in controlling cell growth during certain stages of the cell cycle. *Drosophila* cells lacking this gene grow uncontrollably in culture. Approximately 200 bp of sequence has been determined. The gene has been used as a probe in a low-stringency hybridization to identify a human clone carrying a homologous gene, for which a partial sequence is available. Its function, however, is unknown.

The partial human gene sequence is used as a probe in a Southern analysis of the mouse genome in order to identify a corresponding mouse sequence. The results of the hybridization indicate that the corresponding mouse sequence is a single unique sequence in the mouse genome. It is desired that mutations be identified in the corresponding mouse sequence in order to help to ascertain the human gene function. Therefore, PCR primers are designed. The primer sequences are either based on the partial human sequence which is available or based on the nucleotide sequence of the corresponding mouse sequence, and used according to the

invention to identify mutations in a mouse homolog of the human gene.

That is, DNA samples from the F1 generation of ENU or MMU mutagenised inbred mice are screened using the primers in fSSCP to generate amplimers. Several amplimers containing a mutation are identified via a change in mobility of the amplimer. The mutated organism from which the DNA sample giving rise to the amplimer was obtained is identified, and observed with respect to its phenotype. Because the corresponding *Drosophila* gene is believed to function to control cell growth, phenotypes relating to loss of control of cell growth are observed. Therefore, mice containing mutations in the gene of interest are observed phenotypically for cancer-like conditions.

Alternatively, instead of obtaining DNA samples from the F1 generation of a mutagenised mouse, mouse ES cells are mutagenised with ENU or MMU and DNA samples are prepared from each ES clone. These DNA samples are then analyzed for mutations, as above, using PCR primers unique to the sequence of interest. Several mutations in the mouse sequence of interest are identified, and each corresponding ES clone containing a mutation is transferred to a mouse embryo, and the resultant newborn mouse is observed phenotypically for loss of control of cell growth.

The mutants obtained as described above indicate that the gene of interest has a role in the control of cell growth, not only in Drosophila, but also in a mammalian model system. Based on this information, the human gene is completely sequenced, and studied further for function relating to control of cell growth.

EXAMPLE 8

A random region of the mouse genome is sequenced and an open reading frame is found which is predicted to encode a protein with homology to known transcription factors, and Northern

analysis reveals that this sequence recognizes a transcript expressed in fetal mice; therefore, it is not a pseudogene.

In order to determine the gene function, ENU mutagenesis is performed on male mice, these mice are mated and their F1 offspring are screened for the presence of a mutation in this gene without prior observation of phenotypic alterations resulting from the mutation, all according to the methods described above. Mutations are detected in 14 mice, but not all of these mutations will result in an altered phenotype as some are silent. Upon subsequent phenotypic analysis, some of these mice are found to be uncoordinated and to perform at below normal levels in standard tests of learning and memory, such as a maze test. Post-mortem neurological examination reveals morphological defects in the brain. Developmental defects resulting in, for example, mental deficiency, varying in type and degree, in humans are the focus of many clinical research programs. It is decided that the human homologue of this gene may be linked to a genetic form of mental retardation and should be identified.

The predicted protein-coding sequence of the gene is examined to identify those regions most highly conserved between it and related transcription-factor-encoding genes. These regions serve as the basis for the design of oligonucleotide primers that will be used in a PCR reaction to identify and amplify the corresponding gene from human genomic DNA. The sequence of these oligonucleotide primers matches exactly that of the mouse gene, where a match is required to preserve the coding capacity of the primer; however, the sequence is randomized at positions, such as third positions of codons, where multiple different bases could still preserve coding of a given amino acid (*ie.* degenerate probes). Given an evolutionary distance of approximately 80 million years between mouse and human species (Mitchell *et al.*, 1992, *Nature 359*: 528-531), reaction conditions are routinely determined, and are as follows: 50 pmol of primer is used with 1 μg of human genomic DNA in 7.5 mM $MgCl_2$, 0.2 mM dNTP,

2.5 units Taq polymerase (Perkin Elmer-Cetus) in the manufacturer's buffer. Samples are overlaid with mineral oil, denatured at 94°C for 5 minutes, then amplified over 35 cycles of 94°C for 1 minute, 53°C for 1 minute and 72°C for 90 seconds. Following the last cycle, a final extension is carried out for 7 minutes at 72°C.

The fragment isolated from human DNA is subsequently sequenced and used in further studies to determine its involvement, if any, in clinical risk for known genetic forms of mental retardation in humans.

Therefore, starting with a random coding sequence: mice are identified which carried mutant copies of the gene; a phenotype associated with these mutations is identified, which could not have been predicted from the sequence alone and which was not previously known to be associated with the gene; further mice, both heterozygous and homozygous, carrying mutant copies of the gene are produced by breeding; these mice are used for disease study and modelling; and the human homologue of the gene is identified as a target for further research.

EXAMPLE 9

Albinism I is an autosomal recessive disease characterized by absence of pigment in hair, skin and eyes. Common features in addition to the lack of pigment are reduced visual acuity and photophobia. The disease is known to result from mutation of the tyrosinase gene. Variation in the phenotype in humans has been associated with specific mutations within the gene, and various tyrosinase alleles such as "yellow" and "minimal pigment" exist.

The mouse *Mus musculus* also has a tyrosinase gene and mutations in this gene leads to albino mice. The mouse is a useful tool for studying many human disorders, and the identification of additional alleles of tyrosinase in mouse would allow a greater understanding of the gene function in

mammalian systems.

By screening DNA, a range of tyrosinase mutations can be identified; screening by phenotype will not achieve this. Mutations which cause a albino phenotype identify regions of the protein which are functionally critical; mutations which change an amino acid in tyrosinase but do not disrupt protein function define regions which are less important. If DNA from a mutagenised population is screened, a larger range of mutations will be identified than from a population whose polymorphism arises only from spontaneous mutations. For instance, ENU can be expected to introduce mutations causing a change in phenotype at a frequency 80 times greater than natural spontaneous mutation.

Treatment of male mice with ENU mutates the sperm producing cells. To fix these mutations into the somatic tissue and germ cells of a mouse, ENU treated males are mated to non-mutagenised females. The resulting F1 offspring will be heterozygous, carrying mutations contributed by the mutagenised father. Albinism I is a recessive disease, requiring that both alleles have protein-disrupting mutations to result in the albino phenotype. The maternal allele, being normal, will provide appropriate function so none of the F1 will be albino as a result of mutation of the paternal tyrosinase gene.

To identify those members of the F1 population carrying a mutation in a tyrosinase allele, the DNA can be screened using fSSCP. PCR primers suitable for generating tyrosinase probes are listed in Table 1. The primers listed will generate probes covering 98% of the coding region of the tyrosinase gene (Table 2).

Table 2 also shows how labelling with different coloured fluorochromes means that multiple coloured PCR products and multiple products of the same colour but of different size can be run in a single lane of an fSSCP gel. For example,

primer pairs A (yellow), C (green), D (blue; 306 bp product) and J (blue, 206 bp product) can be used to probe and amplify DNA from an animal, and the products pooled and run in a single lane. In this example, 1042 bp, corresponding to 65% of the protein coding region of the tyrosinase gene, are examined in a single lane of a gel. By running gels with many lanes, for example 66 lanes per gel, and running many gels, for example 250 gels, 10000 ENU F1 mice can be tested for mutation in the tyrosinase gene. In this way, multiple independent mutations in the tyrosinase gene can be discovered.

As albinism I is recessive, the F1 mice with tyrosinase mutations will not be albino. To study the phenotype due to the mutations identified, it is necessary to first cross each F1 mutant mouse to a normal mouse to generate additional mice heterozygous for the mutation. Half of these F2 offspring will be heterozygotic for mutations in the tyrosinase gene and these can easily be identified. If these F2 heterozygotes are bred together, one quarter of the offspring will be homozygous for the mutated tyrosinase gene and may exhibit an altered phenotype (*eg.* different fur colour). The specific mutations in the F2 mice can be determined and the relative effect of various mutations can be compared.

EXAMPLE 10

The *SRY* gene is the dominant inducer of male development found on the Y chromosome. Mutations in the conserved DNA binding domain of SRY in humans have been found which disable protein function and result in individuals with a Y chromosome who develop as females; "sex reversal" has once been observed due to a mis-sense mutation outside the DNA binding domain.

Mice with a Y chromosome carrying a deletion of *Sry* develop as females. However, sex reversed mice appear to the eye as normal females and are not easily spotted. In the absence of DNA tests for the presence of *Sry* reversal is not noted and

no *Sry* mutations are known. It is presumed, however, that mutations within *Sry* will lead to sex reversal in mice, although the types and location of such mutations are not known.

In humans, some *SRY* mutations have variable penetrance. They sometimes do not cause sex reversal but in subsequent generations the Y chromosome bearing offspring with mutant *SRY* develop as females. It is unknown if such mutations exist in mice. It would be valuable to identify such mutations in mice to study sex determination and development abnormalities which have a parallel in humans. These "conditional" mutations cannot be discovered in males by phenotypic assays as there is no testable phenotype in those individuals. The mutation needs to be known and studied in subsequent generations, including studies which transfer the mutated gene to other genetic backgrounds to test the effect of modifier alleles which may be present.

Mutations in the DNA of mice may be generated and discovered by the invention as follows. A population of male mice are treated with ENU to provide a source of mutant *Sry* and a heterozygotic F1 generation is produced. fSSCP is utilised to identify those members of the F1 population carrying *Sry* mutations. Primers for examining the entire protein coding region of *Sry* by fSSCP are shown in Table 3 and details of the position of the primers in *Sry* are shown in Table 4.

Animals carrying mutations in *Sry* can be examined for sexual phenotype, and mutations in the gene causing sex reversal identified. The location of the mutations in the mouse *Sry* gene can be compared to those known in humans to be sex reversing. *Sry* mutations in phenotypic males (*ie.* those which are not sex reversed) can be tested for variable penetrance by breeding to females of the same strain and of different strains. The offspring are tested for mutated *Sry* and the sexual phenotype examined to find if the mutation being studied causes sex reversal. In this way, animal models of

human sex reversal can be derived.

EXAMPLE 11

PAX6 mutations lead to a variety of anterior segment malformations most commonly characterized by eye development defects broadly described as aniridia. The phenotype is panocular and variable, with features ranging from a readily visible, nearly complete absence of the iris, to small slit-like defects in the anterior layer seen only with a slit-lamp. Mice with *Pax-6* mutations display similar phenotypes.

The disease is dominant and apparently results from haploinsufficiency: the loss of function of one allele in the presence of a normal allele. Variability in the phenotype makes ascertainment of all eye abnormalities associated with *Pax-6* difficult. In examination for eye abnormalities, subtle phenotypes resulting from mutation of *Pax-6* can be missed. It is therefore useful first to identify individual mice carrying *Pax-6* mutations and then to look carefully for associated phenotypes.

Such mice can be identified by the invention by screening mice which carry ENU induced mutations. F1 mice which are the progeny of ENU mutagenised fathers and normal mothers are screened by fSSCP to detect heterozygous mutations in the PAX6 gene. Appropriate primers for screening a portion of the Pax-6 gene are described in Tables 5 & 6. Mice which are identified as carrying mutations in *Pax-6* are thoroughly and carefully phenotypically tested for eye abnormalities, including anterior segment anomalies. In this way the phenotypic spectrum associated with *Pax-6* mutations in mice is identified and mouse models for varying severities of aniridia in humans are identified.

Other embodiments

Other embodiments will be evident to those of skill in the art. It should be understood that the foregoing detailed

description is provided for clarity only and is merely exemplary. The spirit and scope of the present invention are not limited to the above examples, but are encompassed by the appended claims.

Industrial applicability

The invention is useful in the discovery and characterization of a gene or genes of interest, with an aim toward the development of therapeutic agents or therapeutic targets for treating human disease or for treating animal diseases. The invention is also useful for DNA mutation screening of a class or family of genes, rather than single genes. The invention provides a rapid assay for the identification of mutant genes and overcomes one of the major obstacles toward the functional analysis of gene function, *ie.* the need for a phenotypic screen to identify a mutation in a gene of interest.

Screening methods of the invention are thus useful for determining the function of a gene for which a function is unknown; for determining the range of phenotypes associated with different mutations in a gene for which a function may or may not have been known; and to identify specified mutations in a gene of interest.

The invention results in the production of organisms which can be used in a number of ways for drug discovery. For example, a mouse containing a mutation in a gene of interest may provide a model for the study and treatment of the disease state; cells derived from the mutant mouse will allow *in vitro* assessment of drug activity; and interbreeding of mutant mice which each have a mutation in a gene involved in a polygenic disorder will allow investigation of gene interactions in the overall phenotype.

The contents of citations mentioned above are fully incorporated herein by reference.

TABLE LEGENDS

Tables 1, 3 & 5

PCR primer sequences for use in fSSCP of the tyrosinase, *Sry* and *Pax-6* genes. "Amplimer name" represents the PCR product amplified by the corresponding pair of forward & reverse PCR primers ("Oligo name"). "Sequencing tail sequence" is the sequence incorporated onto the end of each primer. It is used to prime sequencing reactions to determine the DNA sequence of the fragment amplified by the primers and is not specific to the gene. "Gene Specific Annealing Sequence" is the sequence in each primer which is specific to the gene. "Primer Label Colour" is the colour of fluorescent label attached to each primer. "Gene Tm" is the temperature at which one-half of the primer is dissociated from the specific genomic sequence. This serves as a guide to PCR annealing temperatures.

Tables 2, 4 & 6

"Exon #" indicates the exon number in the gene corresponding to the position of the "Primer Pair". "Primer Pair SSCP Coverage" indicates the amount and position of sequence tested in SSCP using the given primer pair. "Total SSCP Coverage of ORF" indicates the cumulative coverage of the gene being tested, taking into account amplimer overlap.

Table 1

Tyrosinase Gene SSCP Primers

Primer Pair	Amplimer	Oligo Name	Sequencing Tail Sequence	Gene Specific Annealing Sequence	Primer Length	Primer Label Color	Gene Tm	Product Size (bp)
A	cMm100	cMm100f	GTAAAACGACGGCCAGT	TCATTAACCTATTGGTGCAG	37	Yellow	53.8	287
	cMm100	cMm100r	GGAAACAGCTATGACCAT	GGAACTGAGGTCCAGATGGTG	39	Yellow	58.1	
B	cMm120	cMm120f	GTAAAACGACGGCCAGT	CTTTCAGGCAGAGGTTCCTG	37	Blue	60	184
	cMm120	cMm120r	GGAAACAGCTATGACCAT	GTCTCTGTACAATTTGGGCCC	39	Blue	60.1	
C	cMm140	cMm140f	GTAAAACGACGGCCAGT	AATAGGACCTGCCAGTGCTC	37	Green	59.3	243
	cMm140	cMm140r	GGAAACAGCTATGACCAT	GGCCTGTGGGGATGACATAGA	39	Green	60.9	
D	cMm160	cMm160f	GTAAAACGACGGCCAGT	TTGAGTGTCTCCGAAAAGAA	37	Blue	56	306
	cMm160	cMm160r	GGAAACAGCTATGACCAT	GCAAGAAAAGTCTGTGCCAAG	39	Blue	56.1	
E	cMm180	cMm180f	GTAAAACGACGGCCAGT	GATTTTGCCCATGAAGCACC	37	Green	62.7	271
	cMm180	cMm180r	GGAAACAGCTATGACCAT	GTGCATCTTACCTGCCAGGAG	39	Green	61.3	
F	cMm200	cMm200f	GTAAAACGACGGCCAGT	TCACTTTAACATCAAATTGTT	38	Yellow	50.1	292
	cMm200	cMm200r	GGAAACAGCTATGACCAT	GTTATATTACCTTCCAGTGTGTT	41	Yellow	50.9	
G	cMm210	cMm210f	TGTAAAACGACGGCCAGT	ACATCAAATTGTTTTTCACCAG	40	Yellow	56.7	284
	cMm210	cMm210r	CAGGAAACAGCTATGACCAT	ATATTACCTTCCAGTGTGTTTCT	43	Yellow	53.8	
H	cMm220	cMm220f	GTAAAACGACGGCCAGT	TTTAATTTCCCTTTATTCAAC	38	Blue	50.8	217
	cMm220	cMm220r	GGAAACAGCTATGACCAT	GGTCAACCAACCTGTCCACAA	39	Blue	59.4	
I	cMm240	cMm240f	GTAAAACGACGGCCAGT	TCCTGACTCTGAGTAACCCT	37	Yellow	53.8	259
	cMm240	cMm240r	GGAAACAGCTATGACCAT	GTGAGCTTTACCTGACTCTTG	39	Yellow	53.2	
J	cMm260	cMm260f	GTAAAACGACGGCCAGT	TCTGTGAAAACAGCTTGATC	37	Blue	53.9	206
	cMm260	cMm260r	GGAAACAGCTATGACCAT	GAGGCATAGCCTACTGCTAAG	39	Blue	54.2	
K	cMm280	cMm280f	GTAAAACGACGGCCAGT	ACTGGTGGGAGGCTGTTATTG	37	Green	57.7	233
	cMm280	cMm280r	GGAAACAGCTATGACCAT	GGAGGTAAAACCTTTCAGTCC	39	Green	53	

Table 2

Tyrosinase Gene SSCP Summary

Exon #	Exon Size	Primer Pair	Primer Color	Amplimer Size	Primer Pair SSCP Coverage	Sequence Tested per Primer Pair	Total SSCP Coverage of ORF
1	818 bp	cMm100f/r	Yellow	287 bp	62 - 248	186 bp	186 bp
1	818 bp	cMm120f/r	Blue	184 bp	226 - 384	158 bp	322 bp
1	818 bp	cMm140f/r	Green	243 bp	337 - 503	166 bp	441 bp
1	818 bp	cMm160f/r	Blue	306 bp	457 - 686	229 bp	229 bp
1	818 bp	cMm180f/r	Green	271 bp	676 -870	194 bp	808 bp
2	213 bp	cMm200f/r	Yellow	292 bp	881 - 1085	204 bp	1037 bp
2	213 bp	cMm210f/r	Yellow	284 bp	881 - 1081	200 bp	1008 bp
3	147 bp	cMm220f/r	Blue	217 bp	1098 -1236	138 bp	1146 bp
4	181 bp	cMm240f/r	Yellow	259 bp	1246 - 1417	171 bp	1317 bp
5	235 bp	cMm260f/r	Blue	206 bp	1428 - 1543	115 bp	1432 bp
5	235 bp	cMm280f/r	Yellow	233 bp	1527 - 1663	136 bp	1568 bp
Open Reading Frame:	1594 bp					Total SSCP Coverage:	1568 bp

Table 3

Sry Gene SSCP Primers

Primer Pair	Amplimer	Oligo Name	Sequencing Tail Sequence	Gene Specific Annealing Sequence	Primer Length	Primer Color Label Color	Gene Tm	Product Size (bp)
A	SryMm100	SryMm100f	TGTAAAACGACGGCCAGT	CAAGTTTTGGGACTGGTGAC	38	Blue	58.0	220
	SryMm100	SryMm100r	CAGGAAACAGCTATGACCAT	GGCTTCTGTAAGGCTTTTCC	40	Blue	58.1	
B	SryMm120	SryMm120f	TGTAAAACGACGGCCAGT	CAAGTTGGCCCAGCAGAATC	38	Yellow	63.0	263
	SryMm120	SryMm120r	CAGGAAACAGCTATGACCAT	GTTGAGGCAACTGCAGGCTG	40	Yellow	64.4	
C	SryMm140	SryMm140f	TGTAAAACGACGGCCAGT	GAGGGCTAAAGTGTCACAGAGGAG	42	Yellow	62.8	304
	SryMm140	SryMm140r	CAGGAAACAGCTATGACCAT	CATAGAACTGCTGTTGCTGCTGG	43	Yellow	64.4	
D	SryMm160	SryMm160f	TGTAAAACGACGGCCAGT	CAGCAGCAGCAGTTCCATAACC	40	Blue	64.4	223
	SryMm160	SryMm160r	CAGGAAACAGCTATGACCAT	GCTCCTGGTGGTGGTGGTG	39	Blue	65.1	
E	SryMm180	SryMm180f	TGTAAAACGACGGCCAGT	CAGCAGAAGCAGCAGTTT	36	Green	55.2	248
	SryMm180	SryMm180r	CAGGAAACAGCTATGACCAT	TTGGTGGTGGTGGTGGTC	38	Green	61.3	
F	SryMm200	SryMm200f	TGTAAAACGACGGCCAGT	ACCACCACCACCAGGAGC	36	Green	62.1	261
	SryMm200	SryMm200r	CAGGAAACAGCTATGACCAT	TGATGCTGCTGCTGCTGG	38	Green	63.6	
G	SryMm220	SryMm220f	TGTAAAACGACGGCCAGT	CACCACCACCACCAAC	34	Blue	54.3	180
	SryMm220	SryMm220r	CAGGAAACAGCTATGACCAT	GTTGCTGCTGGGGGTG	36	Blue	59.7	
H	SryMm240	SryMm240f	TGTAAAACGACGGCCAGT	CATCAGTTCCATGACCACCC	38	Green	61.2	243
	SryMm240	SryMm240r	CAGGAAACAGCTATGACCAT	GGTCATGGAACTGCTGTTGC	40	Green	61.7	
I	SryMm260	SryMm260f	TGTAAAACGACGGCCAGT	GAAGCAGCAGTTCCATGACC	38	Yellow	60.8	282
	SryMm260	SryMm260r	CAGGAAACAGCTATGACCAT	CATAGCAAGGGGGAGTGTTG	40	Yellow	60.5	

Table 4

Sry Gene SSCP Summary

Exon #	Exon Size	Primer Pair	Primer Color	Amplimer Size	Primer Pair SSCP Coverage	Sequence Tested per Primer Pair	Total SSCP Coverage of ORF
1	1187 bp	sryMm100f/r	Blue	220 bp	51 - 177	126 bp	126 bp
1	1187 bp	sryMm120f/r	Yellow	263 bp	133 - 317	184 bp	266 bp
1	1187 bp	sryMm140f/r	Yellow	304 bp	308 - 526	218 bp	475 bp
1	1187 bp	sryMm160f/r	Blue	223 bp	500 - 665	165 bp	614 bp
1	1187 bp	sryMm180f/r	Green	248 bp	628 - 800	172 bp	749 bp
1	1187 bp	sryMm200f/r	Green	261 bp	685 - 870	185 bp	819 bp
1	1187 bp	sryMm220f/r	Blue	180 bp	820 - 929	109 bp	878 bp
1	1187 bp	sryMm240f/r	Green	243 bp	903 - 1069	166 bp	1018 bp
1	1187 bp	sryMm260f/r	Yellow	282 bp	1060 - 1238	178 bp	1187 bp
Open Reading Frame:	1187 bp					Total SSCP Coverage:	1187 bp

Table 5

Pax-6 gene SSCP primers

Primer Pair	Amplimer	Oligo Name	Sequencing Tail Sequence	Gene Specific Annealing Sequence	Primer Length	Primer Label Color	Gene Tm	Product Size (bp)
A	Pax6Mm121	Pax6Mm121f	TGTAAAACGACGGCCAGT	GTCACAGCGGAGTGAATCAG	38	Blue	59.4	168 bp
	Pax6Mm121	Pax6Mm121r	CAGGAAACAGCTATGACCAT	TGCAGAATTCGGGAAATGTC	40	Blue	61	
B	Pax6Mm161	Pax6Mm161f	TGTAAAACGACGGCCAGT	GTATCCAACGGTTGTGTGAG	38	Yellow	56.4	254 bp
	Pax6Mm161	Pax6Mm161r	CAGGAAACAGCTATGACCAT	ACTGGGTATGTTATCGTTGG	40	Yellow	55	
C	Pax6Mm180	Pax6Mm180f	TGTAAAACGACGGCCAGT	GTGTCATCAATAAACAGAGTTCTTC	43	Yellow	56.2	200 bp
	Pax6Mm180	Pax6Mm180r	CAGGAAACAGCTATGACCAT	CGTGGGTTGCCCTGGTAC	38	Yellow	62.4	
D	Pax6Mm200	Pax6Mm200f	TGTAAAACGACGGCCAGT	ATGGCTGCCAGCAACAGG	36	Green	63.4	197 bp
	Pax6Mm200	Pax6Mm200r	CAGGAAACAGCTATGACCAT	CTTTCTCCAGAGCCTCAATCTGC	43	Green	63.9	
E	Pax6Mm220	Pax6Mm220f	TGTAAAACGACGGCCAGT	GTATGGTTTTCTAATCGAAGGGC	41	Green	60.2	189 bp
	Pax6Mm220	Pax6Mm220r	CAGGAAACAGCTATGACCAT	CAGGTGTGGTGGGCTGTG	38	Green	62.3	
F	Pax6Mm240	Pax6Mm240f	TGTAAAACGACGGCCAGT	CCCCCAGTCCCCAGTCAG	36	Blue	64	180 bp
	Pax6Mm240	Pax6Mm240r	CAGGAAACAGCTATGACCAT	CTGTTGAAGTGGTCCCCGAG	40	Blue	63	
G	Pax6Mm260	Pax6Mm260f	TGTAAAACGACGGCCAGT	GACTCATTTCACCTGGAGTG	38	Blue	55.5	297 bp
	Pax6Mm260	Pax6Mm260r	CAGGAAACAGCTATGACCAT	AAGGTCCTTGGTTCTAGTCC	40	Blue	54.9	

Table 6

Pax-6 Gene SSCP Summary

Exon #	Exon Size	Primer Pair	Primer Color	Amplimer Size	Primer Pair SSCP Coverage	Sequence Tested per Primer Pair	Total SSCP Coverage of ORF
5	130 bp	paxMm100f/r	Blue	168 bp	191 - 282	91 bp	91 bp
6	215 bp	paxMm120f/r	Yellow	254 bp	366 - 541	175 bp	266 bp
7	165 bp	paxMm140f/r	Yellow	200 bp	587 - 705	118 bp	384 bp
8	158 bp	paxMm160f/r	Green	197 bp	746 - 863	117 bp	501 bp
10	150 bp	paxMm180f/r	Green	189 bp	993 - 1102	109 bp	610 bp
12	150 bp	paxMm200f/r	Blue	189 bp	1255 - 1367	112 bp	722 bp
13	85 bp	paxMm220f/r	Blue	297 bp	1407 - 1473	66 bp	788 bp
Open Reading Frame:	2098 bp					Total SSCP Coverage:	788 bp

CLAIMS

1. A method of identifying the presence of a mutation in a gene of interest in an organism, comprising testing a nucleic acid sample from a mutated organism for a mutation in a gene of interest without the prior observation of a phenotypic alteration in said mutated organism resulting from said mutation.

2. A method of identifying the presence of a mutation in two or more genes of interest in an organism, the method comprising the steps of

a) providing a nucleic acid sample from a given tissue of a mutated organism, wherein said mutated organism contains about 1 mutation in 1000 genes.

b) providing plural nucleic acid probes, each said probe being unique to a given gene of interest,

c) testing said nucleic acid sample for the presence of a mutation in a said gene of interest.

3. The method of claim 1 or claim 2 further comprising the initial step of mutagenising an organism so as to produce said mutated organism.

4. A method of identifying the presence of a mutation in a gene of interest in an organism, comprising mutagenising a plurality of the same organism to produce mutated organisms; and testing a mixture of pooled nucleic acid samples said mixture comprising a plurality of nucleic acid samples from a corresponding plurality of mutated organisms for a mutation in a gene of interest without the prior observation of a phenotypic alteration in said mutated organisms resulting from said mutation.

5. The method of any of claims 1-4 wherein each mutated organism contains about 1 mutation in every 10000 genes to 1 mutation in every 1000 genes.

6. The method of any of claims 3, 4 or 5 wherein said mutagenising step comprises inducing a genetic mutation into an organism at an average frequency of 1/1000 to 1/10000 organisms.

7. The method of claim 6 wherein said mutagenising step comprises exposing said organism to an alkylating agent, such as ENU or MNU.

8. The method of claim 6 wherein the mutation is a single base pair mutation.

9. The method of any preceding claim wherein said plurality of mutated organisms comprise about 1000-10000 organisms.

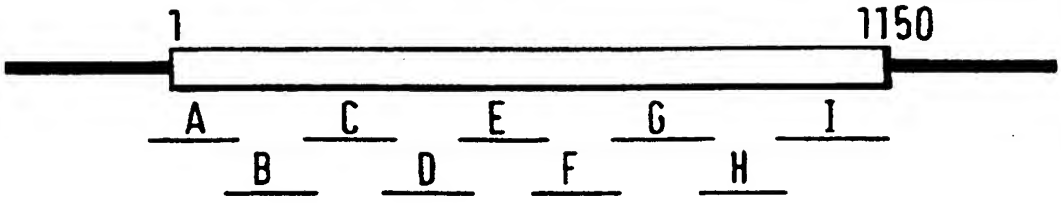
10. A method of identifying a gene of interest in a first organism, comprising in order the steps of:

a) identifying in a second organism of a species different from said first organism a mutation in a gene of interest in a nucleic acid sample from a mutated second organism without the prior observation of a phenotypic alteration in said mutated second organism resulting from said mutation;

b) correlating said mutation with a phenotypic alteration in said mutated second organism resulting from said mutation; and

c) utilizing said gene of interest or sequences thereof from said second organism to identify said gene of interest from said first organism.

FIG. 1



INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/GB 97/01354

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68 A01K67/027 A01K67/033 C12N15/01

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q A01K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MULLINS M. C. ET AL.,: "Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate" CURRENT BIOLOGY, vol. 4, no. 3, - 3 March 1994 pages 189-202, XP002042326 cited in the application see the whole document --- -/--	1-10

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 October 1997	29.10.97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Müller, F

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SIMA A. & SHIMADA A.: "Development of a possible nonmammalian test system for radiation-induced germ-cell muatagenesis using a fish, the japanese medaka (O. laptipes)" PROC. NATL. ACAD. SCI. USA, vol. 88, - March 1991 pages 2545-2549, XP002042327 cited in the application see the whole document ---	1-10
A	DATABASE WPI Section Ch, Week 9402 Derwent Publications Ltd., London, GB; Class B04, AN 94-011017 XP002042328 & JP 05 317 048 A (MATSUDA I) , 3 December 1993 see abstract -----	1-10

3